

# RSD-YOLO: An Improved YOLOv11n-Based Algorithm for Railway Surface Defect Detection

**ChenXin Pan**

Tianjin University of Technology and Education, Tianjin, 300222, China

Email: 2541509878@qq.com

**How to cite this paper:** Pan, C. X. (2026). RSD-YOLO: An improved YOLOv11n-based algorithm for railway surface defect detection. *Academic Journal of Emerging Technologies*, 3(1), 93–102. ISSN Print: 3104-4417, ISSN Online: 3104-4425.

<https://doi.org/10.63313/AJET.9057>

**Published: 2026-05-22**

Copyright © 2026 by author(s) and Erytis Publishing Limited.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



## Abstract

Railway surface defect detection is a critical component of automated track maintenance and safety assurance. To improve the detection capability of lightweight YOLOv11n on rail surface images, this paper proposes RSD-YOLO, a conservative structural enhancement that adds a P2 high-resolution detection branch, replaces selected high-level C3k2 blocks with C2f modules, and introduces P2-SPPF context enhancement. The resulting detector changes the original three-scale P3/P4/P5 output to a four-scale P2/P3/P4/P5 output. Experiments on a four-class rail surface defect dataset show that RSD-YOLO achieves 0.7966 mAP@0.5 and 0.5616 mAP@0.5:0.95 on the test set, improving YOLOv11n by 8.05 and 7.24 percentage points, respectively.

## Keywords

RSD-YOLO; YOLOv11n; Railway Surface Defect Detection; P2 Detection Head; C2f; SPPF

## 1. Introduction

Railway infrastructure safety directly determines the reliability and operational continuity of rail transport systems. Rail surface defects, including spalling, wheel burn, squat, and corrugation, can gradually evolve into serious structural damage if they are not identified and repaired in time. Manual inspection and traditional image-processing methods are vulnerable to subjective judgment, illumination variation, surface contamination, and complex metallic textures. Therefore, an automatic visual inspection method with both high accuracy and real-time performance is required for practical railway maintenance.

Deep learning-based object detection has become an important technical route for industrial defect inspection. Two-stage detectors provide strong localization performance but usually cannot satisfy the speed requirements of on-board inspection. One-stage detectors represented by the YOLO series directly regress bounding boxes and categories from images, offering a favorable balance between

accuracy and efficiency. YOLOv11 further improves feature extraction and prediction efficiency, and the lightweight YOLOv11n variant is suitable for deployment scenarios.

However, railway surface defects have strong scale diversity and weak visual boundaries. Spalling and squat may appear as small localized regions, wheel burn often has diffuse thermal boundaries, and corrugation appears as periodic shallow patterns along the rail surface. The original YOLOv11n detector outputs only P3, P4, and P5 feature scales, which may weaken the representation of fine-grained defect cues.

Therefore, this paper proposes RSD-YOLO, an improved YOLOv11n-based object detection algorithm for railway surface defect detection. The specific improvements are as follows:

- (1) A P2 high-resolution detection branch is added to the neck, changing the final detection output from P3/P4/P5 to P2/P3/P4/P5 and improving sensitivity to small rail defects.
- (2) Selected high-level backbone C3k2 blocks at the P4 and P5 stages are replaced with C2f modules, improving lightweight feature reuse and gradient flow without introducing heavy attention modules.
- (3) An SPPF enhancement unit is introduced on the P2 branch, providing local contextual aggregation before high-resolution detection and reducing false localization caused by rail background texture.

## 2. YOLOv11 Algorithm

YOLOv11 is a new-generation one-stage object detection framework released by Ultralytics. It inherits the end-to-end detection paradigm of the YOLO series and uses a backbone, neck, and decoupled detection head to perform multi-scale object recognition. In the YOLOv11n baseline used in this work, the input image is processed by successive convolutional and C3k2 blocks to generate pyramid features at P3, P4, and P5.

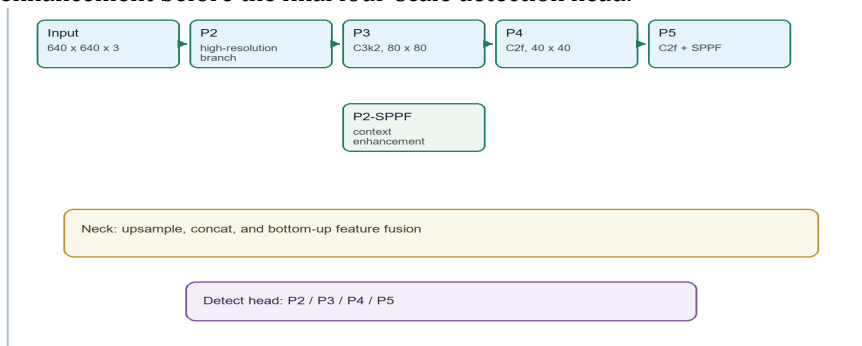
(1) Backbone: The backbone is responsible for hierarchical feature extraction. The baseline configuration uses convolutional layers and C3k2 modules to obtain multi-level semantic features. These features provide the basis for detecting rail defects under complex illumination and background conditions.

(2) Neck: The neck adopts an upsampling, concatenation, and bottom-up fusion structure to combine semantic information from deep layers with spatial information from shallow layers. This fusion mechanism enables the detector to recognize defects at different scales.

(3) Head: The baseline YOLOv11n detection head outputs three scales, namely P3, P4, and P5. Although this design is efficient, small or weakly contrasted rail defects may lose important localization cues after repeated downsampling, which motivates the proposed four-scale detection design.

### 3. RSD-YOLO

This paper improves the YOLOv11n baseline while maintaining efficient inference suitable for railway inspection deployment. Figure 1 shows the overall structure of RSD-YOLO. The improved model adds a P2 detection branch, replaces selected high-level C3k2 blocks with C2f modules, and introduces P2-SPPF context enhancement before the final four-scale detection head.



**Figure 1.** RSD-YOLO Network Structure

#### 3.1. P2 High-Resolution Detection Branch

The first improvement is the introduction of a P2 high-resolution detection branch. In the baseline model, feature fusion stops at P3 before the bottom-up path returns to P4 and P5. RSD-YOLO adds an additional upsampling operation after the P3 fusion layer and concatenates the result with the shallow P2 backbone feature, creating a high-resolution feature map for fine defect localization.

This branch is especially useful for localized or weak rail surface defects whose visual evidence may be partially lost after repeated downsampling. By adding the P2 output to the final Detect layer, the detector receives an additional prediction scale with richer spatial detail. The final detection head therefore changes from three scales to four scales: P2, P3, P4, and P5.

Compared with the original P3/P4/P5 output, the additional P2 layer provides a denser prediction grid. This helps the model preserve boundary information for spalling, squat, and subtle corrugation patterns while keeping the overall network structure lightweight.

The P2 branch is not designed as an isolated shallow detector. It is connected to the subsequent feature-fusion pathway, allowing high-resolution spatial information to participate in both top-down and bottom-up fusion.

#### 3.2. C2f-Based Lightweight High-Level Feature Extraction

The second improvement replaces selected high-level C3k2 blocks with C2f modules. In the improved configuration, the backbone layers corresponding to P4 and P5 feature extraction use C2f blocks with 512 and 1024 channels, respectively. C2f emphasizes feature reuse and gradient flow through a compact split-and-fusion

structure.

For railway surface defect detection, high-level semantic features remain important because defect appearance is influenced by surrounding rail texture, illumination, and longitudinal surface patterns. Using C2f in deeper stages provides a conservative lightweight design that improves feature propagation without adding heavy attention or transformer-style modules.

This modification is consistent with the deployment requirement of railway inspection. It preserves the compact nature of YOLOv11n while strengthening the representation of medium- and large-scale contextual information before it is fused with the P2 high-resolution branch.

### 3.3. P2-SPPF Context Enhancement

The third improvement is the SPPF enhancement unit placed on the P2 branch after concatenating the upsampled P3 feature with the shallow P2 feature. The SPPF module aggregates local multi-scale context through fast spatial pyramid pooling and outputs a compact 128-channel representation.

Rail surface textures are complex, and isolated texture fragments can resemble defects. Small defects may require nearby structural context to avoid false localization. The P2-SPPF unit therefore strengthens the high-resolution branch by adding local context before detection.

The final RSD-YOLO detection layer uses four outputs, Detect(P2, P3, P4, P5), rather than the original Detect(P3, P4, P5). This design improves the ability to detect fine rail defects while preserving real-time inference.

## 4. Dataset and Evaluation Metrics

### 4.1. Experimental Setup and Dataset

All experiments were conducted in a Linux environment with Python 3.8.10, PyTorch 1.11.0+cu113, CUDA 11.3, Ultralytics 8.3.168, OpenCV 4.12.0, and NumPy 1.24.4. The GPU device was an NVIDIA A40. The input size used in the experiment records was 640 x 640, and the batch size shown in the test screenshots was 16.

The dataset contains four categories of railway surface defects: Spalling, Wheel Burn, Squat, and Corrugation. The local dataset contains 1,916 training images, 240 validation images, and 240 test images. A total of 3,950 labeled instances were found, including 1,267 Spalling instances, 454 Wheel Burn instances, 1,568 Squat instances, and 661 Corrugation instances. Table 1 summarizes the defect categories.

**Table 1.** Categorization and descriptions of rail surface defects.

Category	Description
Spalling	Localized material fracture or pitted region with irregular edges; 1,267 labeled instances.
Wheel Burn	Thermally induced surface damage with diffuse and discolored boundaries; 454 labeled instances.

Squat	Rolling-contact fatigue defect appearing as dark indentation or crack-like region; 1,568 labeled instances.
Corrugation	Periodic longitudinal roughness pattern along the rail surface; 661 labeled instances.

Figure 2 shows representative annotated samples from the test set. These samples demonstrate that the four categories differ significantly in morphology, boundary clarity, and scale, which makes multi-scale detection necessary.



**Figure 2.** Sample Dataset Images

## 4.2. Evaluation Metrics

This paper uses Precision, Recall, Average Precision (AP), mAP@0.5, and mAP@0.5:0.95 as evaluation metrics for object detection. Precision measures the proportion of correct positive predictions, while Recall measures the proportion of ground-truth targets that are successfully detected. AP summarizes the precision-recall curve of a single class, and mAP averages AP across all defect categories.

In addition to accuracy, inference speed is reported as frames per second because railway inspection systems require real-time or near-real-time processing. The experimental analysis therefore considers both detection accuracy and computational efficiency.

## 5. Experimental Results and Analysis

### 5.1. Ablation Experiment Results and Analysis

Ablation experiments were conducted to evaluate the contribution of each structural improvement. The baseline is YOLOv11n with three detection scales. The ab01 model adds the P2 detection branch, the ab02 model further replaces selected C3k2 blocks with C2f, and the ab03 model adds the P2-SPPF enhancement unit to form the final RSD-YOLO.

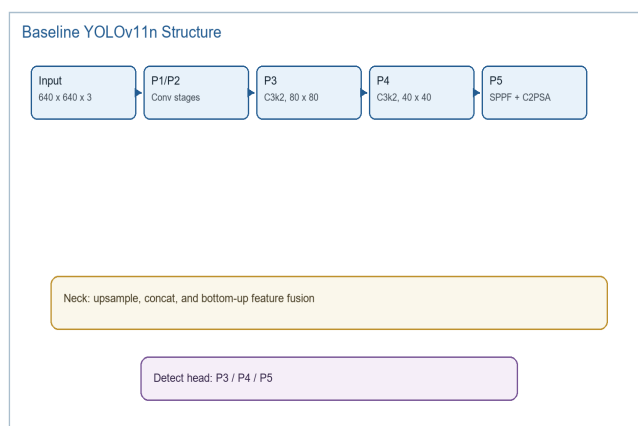
**Table 2.** Ablation Experiment Results Comparison

Model	P2	C2f	P2-SPPF	mAP@0.5	mAP@0.5:0.95	Recall	FPS
-------	----	-----	---------	---------	--------------	--------	-----

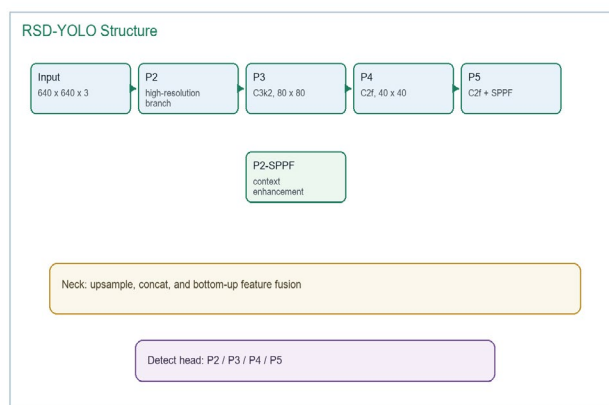
YOLOv11n	x	x	x	0.7161	0.4892	0.7179	99.42
ab01	yes	x	x	0.7799	0.5580	0.8521	90.65
ab02	yes	yes	x	0.7881	0.5695	0.8606	94.34
RSD-YOLO	yes	yes	yes	0.7966	0.5616	0.8375	96.82

As shown in Table 2, introducing the P2 detection branch provides the largest single improvement, increasing mAP@0.5 from 0.7161 to 0.7799. Adding C2f further improves both mAP@0.5 and mAP@0.5:0.95, indicating that lightweight high-level feature reuse benefits defect representation.

The final RSD-YOLO obtains the highest mAP@0.5 of 0.7966 and maintains a high FPS of 96.82. Although ab02 has a slightly higher mAP@0.5:0.95 than the final model, RSD-YOLO provides the best mAP@0.5 and a favorable practical trade-off for rail defect localization.



**Figure 3.** Baseline YOLOv11n Model



**Figure 4.** RSD-YOLO Model

## 5.2. Different Model Comparison Experiment Results and Analysis

To further validate the effectiveness of RSD-YOLO, it is compared with representative lightweight YOLO variants under the same rail surface defect task.

The values for YOLOv8n and YOLOv11n were cross-checked between screenshots and the experiment spreadsheet, while the YOLOv12n row was taken from the spreadsheet record. Table 3 lists the comparison results.

**Table 3.** Different Model Comparison Experiment Results

Model	mAP@0.5	mAP@0.5:0.95	Recall	FPS
YOLOv5n	0.7557	0.4649	0.7487	99.12
YOLOv8n	0.7793	0.5170	0.8063	69.32
YOLOv11n	0.7161	0.4892	0.7179	99.42
YOLOv12n	0.7770	0.5120	-	-
RSD-YOLO	0.7966	0.5616	0.8375	96.82

As shown in Table 3, RSD-YOLO achieves the best mAP@0.5 and mAP@0.5:0.95 among the compared models. Compared with the YOLOv11n baseline, it improves mAP@0.5 by 8.05 percentage points and mAP@0.5:0.95 by 7.24 percentage points. Compared with YOLOv8n, RSD-YOLO improves mAP@0.5 by 1.73 percentage points and mAP@0.5:0.95 by 4.46 percentage points. These results indicate that the proposed high-resolution branch and lightweight feature-extraction modifications are effective for the rail surface defect dataset.

### 5.3. Visualization Comparison Experiment

Considering the complex and variable appearance of rail surface defects, visualization and screenshot comparisons were used to more intuitively verify the detection adaptability of the improved model. Figures 5, 6, and 7 compare the baseline, intermediate ablation models, and the final RSD-YOLO test outputs.

The baseline YOLOv11n model is fast, but its weak performance on corrugation and some low-contrast defects indicates that fine surface texture cues are not sufficiently represented by the original three-scale head. After adding the P2 branch, the detector becomes more sensitive to small and elongated patterns, which substantially improves recall.

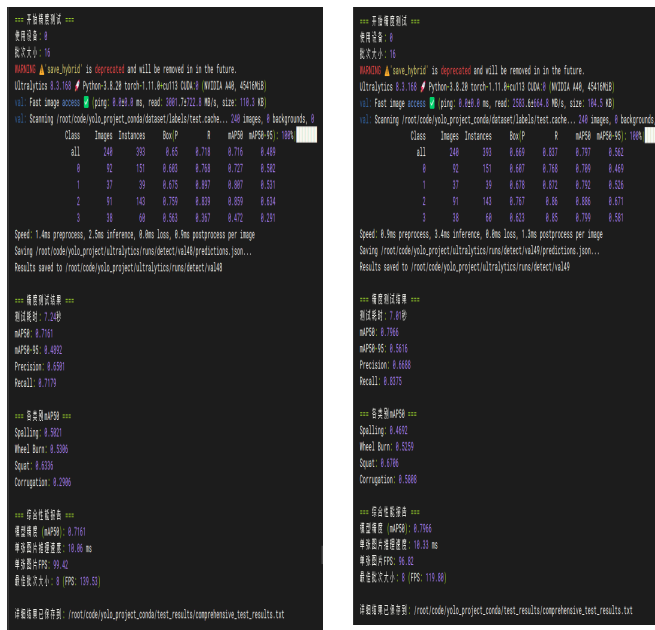


Figure 5. Comparison of YOLOv11n and RSD-YOLO Test Results

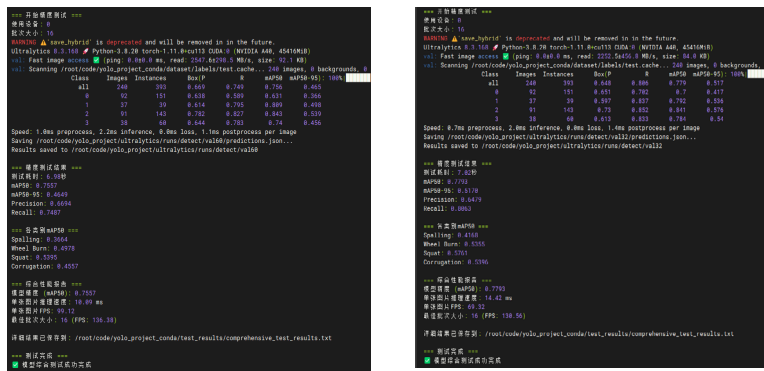


Figure 6. Comparison of YOLOv5n and YOLOv8n Test Results

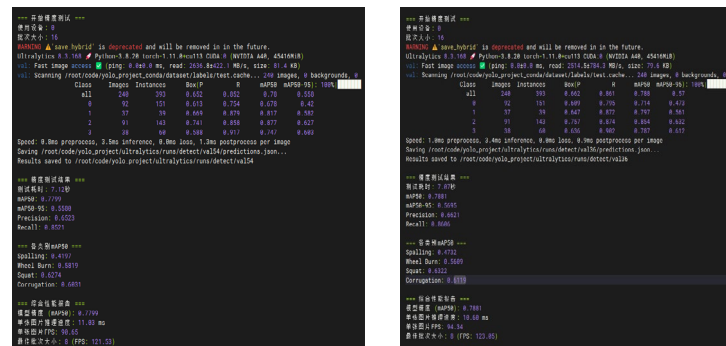


Figure 7. Comparison of ab01 and ab02 Test Results

## 6. Conclusion

This paper presents RSD-YOLO, an improved YOLOv11n-based object detection algorithm specifically designed for railway surface defect detection. Building on the YOLOv11n baseline, RSD-YOLO adds a P2 high-resolution detection branch, replaces selected high-level C3k2 blocks with C2f modules, and introduces P2-SPPF context enhancement.

Experimental evaluations on a four-class railway surface defect dataset demonstrate that RSD-YOLO achieves 0.7966 mAP@0.5, 0.5616 mAP@0.5:0.95, 0.6668 precision, 0.8375 recall, and 96.82 FPS on the test set. Compared with the YOLOv11n baseline, the proposed model improves mAP@0.5 by 8.05 percentage points and mAP@0.5:0.95 by 7.24 percentage points.

Future work will focus on improving robustness under more complex illumination and rail contamination conditions, expanding the dataset with additional field scenarios, and optimizing the model for embedded railway inspection platforms.

## References

- [1] Su, B., Chen, H., & Zhou, Z. (2022). BAF-detector: An efficient CNN-based detector for photovoltaic cell defect detection. *IEEE Transactions on Industrial Electronics*, 69(3), 3161-3171.
- [2] Zhang, Y., Zhang, H., Huang, Q., Han, Y., & Zhao, M. (2024). DsP-YOLO: An anchor-free network with DsPAN for small object detection of multiscale defects. *Expert Systems with Applications*, 238, 122669.
- [3] Hao, X., et al. (2024). An improved YOLOv8 algorithm for rail surface defect detection. *IEEE Access*, 12, 45210-45222.
- [4] Liu, J., Shen, X., Wang, J., Jiang, L., & Zhang, H. (2023). An intelligent defect detection approach based on cascade attention network under complex magnetic flux leakage signals. *IEEE Transactions on Industrial Electronics*, 71(2), 2056-2066.
- [5] Kim, H., et al. (2022). Defect detection of subway tunnels using advanced U-Net network. *Sensors*, 22(6), 2335.
- [6] Li, X., Pu, X., Ling, W. et al. YOLO-SAM an end-to-end framework for efficient real time object detection and segmentation. *Sci Rep* 15, 40854 (2025). <https://doi.org/10.1038/s41598-025-24576-6>
- [7] Han Y, Qi K, Zheng J, et al. Lightweight Cattle Facial Recognition Method Based on Improved YOLOv11. *Smart Agriculture*, 2025, 7(3): 173-184. <https://doi.org/10.12133/j.smartag.SA202502010>
- [8] Xu, X. et al. Exploring image enhancement for salient object detection in low light images. *ACM Trans. Multimed. Comput. Commun. Appl.* 17, 1 - 19 (2021).
- [9] Gao, L., Zhang, J., Yang, C., & Zhou, Y. (2022). Cas-VSwin transformer: A variant swin transformer for surface-defect detection. *Computers in Industry*, 140, 103689.
- [10] Wang Yin, Wang Lide, Qiu Ji. Real-time Enhancement Algorithm for Orbital Dark Light Environment Based on DenseNet Structure[J]. *Journal of Southwest Jiaotong University*, 2022, 57(6): 1349-1357. doi: 10.3969/j.issn.0258-2724.20210199
- [11] Xue R ,Duan J ,Du Z .MPE-DETR: A multiscale pyramid enhancement network for object detection in low-light images[J].*Image and Vision Computing*,2024,150105202-105202.DOI:10.1016/J.IMAVIS.2024.105202.
- [12] Dong, X., Zhang, C., Wang, J., Chen, Y., & Wang, D. (2024). Real-time detection of surface cracking defects for large-sized stamped parts. *Computers in Industry*, 158, 104105.

- [13] Chen, Z., Yang, C., Li, Q., Zhao, F., Zha, Z. J., & Wu, F. (2024). A Dual-Task Branch Alignment Detector for Object Detection. *IEEE Transactions on Image Processing*, 33, 1642-1655.